MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2017/2018

DSP5018 – STATISTICS AND PROBABILITY

(*for diploma students only*)

3 MARCH 2018

02.30 pm – 4.30 pm

(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of **4** pages excluding cover page and appendix.

2. Attempt **ALL FOUR (4)** questions. The marks distribution for each question is given.

3. Write your answers in the answer booklet provided.

4. Key formulae are given in the Appendix.

Question 1

a) The following stem and leave diagram shows the marks obtained in Business Mathematics final exam by a group of 20 students.

STEM	LEAF							
3	4	9						
4	3							
5	4							
6	7							
7	0	5	6					
8	1	2	4	5	6	8	9	9
9	0	6	6	9				

Key: 7|0 means 70

i)   Based on the diagram, list down the 5 number summary of the distribution.   (6.5 marks)

ii) Calculate the interquartile range, and identify the lower and upper limit of the distribution.

(3 marks)

iii) Identify any outliers and draw a boxplot to represent the above data.   (4 marks)

b) Result of a study on 31 lecturers in MMU Melaka regarding their weekly intake of vitamins, in grams, are shown in the following histogram.
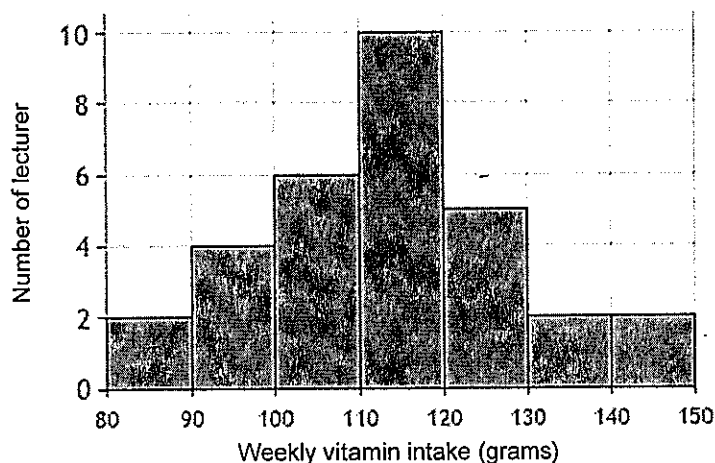
i) Reconstruct and complete the grouped data table below based on the data in the histogram.

(6 marks)

Class/Vitamin Intake (grams)	Frequency	Relative frequency	Midpoint
80<90			
Total			

ii) From table in part (i), calculate the standard deviation of the weekly intake of vitamin. Give your answer in four decimal places. (5.5 marks)

[TOTAL 25 MARKS]

Question 2

a) A group of students study the behavior of customers at a fast food restaurant. As it turns out, 60% of their business operation is via the drive-through window (D), and 40% is via counter-service (C). Data from the registers reveal that 70% of the transactions at the drive-through are via credit/debit cards (M), while 80% of the transactions at the counter are via credit/debit cards and the other transactions are via cash (N).

i) Construct a tree diagram to represent the above information. (3 marks)

ii) What is the probability of a random customer to make order via drive-through and uses credit card? (2 marks)

iii) What is the probability that a random transaction uses cash? (3 marks)

b) In a group of 50 people waiting at a bus stop one morning, 30 people had an umbrella, 20 people had a raincoat and 12 people had both an umbrella and a raincoat.

i) Draw a Venn diagram to represent the above information, using "U" for the event "has an umbrella" and "R" for the event "has a raincoat". (3 marks)

ii) If a person was randomly picked from the bus stop, what is the probability that the person had a raincoat only? (1 mark)

iii) Are the event U and R mutually exclusive? Explain your answer. (2 marks)

Continued...

c) The table below shows the results of a survey regarding how MMU students get to the campus each day.

	Male	Female	TOTAL
Car	8	v	11
Others	w	5	x
TOTAL	y	z	23

    i) Complete the table by finding the values of v, w, x, y and z. (5 marks)

    ii) If one student is chosen at random, find the probability to choose a male student, given that he gets to MMU campus by car. (3 marks)

    iii) If a student gets to MMU campus by bus, find the probability that the student is a female student. (3 marks)

[TOTAL 25 MARKS]

Question 3

	Uber Rental	Grab Rental
Sample mean	RM 35	RM 40
Sample standard deviation	RM 17.50	RM 23.50

a) The table above shows the survey of 100 respondents for each car rental services in Melaka area. It is claimed that there is a difference in the average price between Uber Rental and Grab Rental. At 10% significance level, can we conclude that the claim is true? (10 marks)

b) According to Mekdi Food & Restaurants, it was reported that more than 85% of students bought the Prosperity Burger during Chinese New Year. A random sample of 250 students was selected and 217 claimed that they bought the Prosperity Burger. Test at 10% of significance level whether to reject or accept the claim. (11 marks)

c) For each of the following statements, write (T) for a true statement and (F) for a false statement. (4 marks)

    i) Hypothesis testing is one of the methods commonly used for making decisions or judgments.

    ii) A hypothesis test involves two hypotheses; the null hypothesis and the alternative hypothesis.

    iii) There are two possible choices for the alternative hypothesis.

    iv) A hypothesis test is called a one-tailed test if it is either left tailed or right tailed.

[TOTAL 25 MARKS]

Continued...

Question 4

Malaysian Youth Club organizes a street collection for a mental health charity. The location takes place in a large city on a particular Sunday. 10 volunteers each holding donation box, stand in busy places and ask passers-by for donations. The table below shows the time they spent together, x minutes, and the collection amounts, y, to the nearest RM.

Collector	A	B	C	D	E	F	G	H	I	J
x	45	60	75	90	110	65	85	100	120	140
y	120	320	280	300	288	150	170	295	410	385

a) Given $\sum y^2 = 822894, \sum x^2 = 86900, \sum xy = 261080.$ Compute the value of SS_{xx}, SS_{yy} and SS_{xy}. (3 marks)

b) Calculate the correlation coefficient, r and interpret the answer. (3 marks)

c) Compute the regression equation $\hat{y} = \beta_0 + \beta_1 x$. (4 marks)

d) Find the sum of squares of SST, SSR and SSE. (3 marks)

e) Calculate the standard error of the estimate, s_e. (2 marks)

f) At 5% significance level, do the data provide sufficient evidence to conclude that time is useful as predictor of amount? (8 marks)

g) Predict the amount of donations collected by volunteers in 160 minutes. (2 marks)

[TOTAL 25 MARKS]

End of page.

APPENDIX – KEY FORMULA

	Ungrouped Data	**Grouped Data**
Mean	$$\bar{x} = \frac{\sum x_i}{n}$$ where n : sample size	$$\bar{x} = \frac{\sum m_i f_i}{\sum f_i}$$ where m : class midpoint $\quad f$: class frequency
Variance	$$s^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right]$$ where n : sample size	$$s^2 = \frac{1}{\left(\sum f_i\right)-1}\left[\sum m_i^2 f_i - \frac{\left(\sum m_i f_i\right)^2}{\sum f_i}\right]$$ where m : class midpoint $\quad f$: class frequency

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $\quad Z = \dfrac{x - \mu}{\sigma}$

HYPOTHESIS TESTING

Population	Mean	Proportion
1	$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$	$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \quad \text{where} \quad \hat{p} = \frac{x}{n}$$
2	$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$	$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}_p(1-\hat{p}_p)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$ $$\text{where} \quad \hat{p}_p = \frac{x_1 + x_2}{n_1 + n_2}$$

Continued…

LINEAR REGRESSION ANALYSIS

- **Least square regression equation,** $\hat{y} = \beta_0 + \beta_1 x$ where $\beta_1 = \dfrac{S_{xy}}{S_{xx}}$ and $\beta_0 = \bar{y} - \beta_1 \bar{x}$

- **Correlation coefficient,** $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ where

$S_{xx} = \sum x^2 - \dfrac{\left(\sum x\right)^2}{n}$	$S_{yy} = \sum y^2 - \dfrac{\left(\sum y\right)^2}{n}$	$S_{xy} = \sum xy - \dfrac{\left(\sum x\right)\left(\sum y\right)}{n}$

- **Sum of Square**

$SST = SSR + SSE$	$SST = S_{yy}$	$SSR = \dfrac{\left(S_{xy}\right)^2}{S_{xx}}$

- **Test of Significance for Regression Slope**

Standard Error, $s_e = \sqrt{\dfrac{SSE}{n-2}}$	Regression t-Test $t = \dfrac{\beta_1}{s_e / \sqrt{Sxx}}$	Correlation t-Test $t = r \cdot \sqrt{\dfrac{n-2}{1-r^2}}$

End of page.